

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
31. Mai 2001 (31.05.2001)

PCT

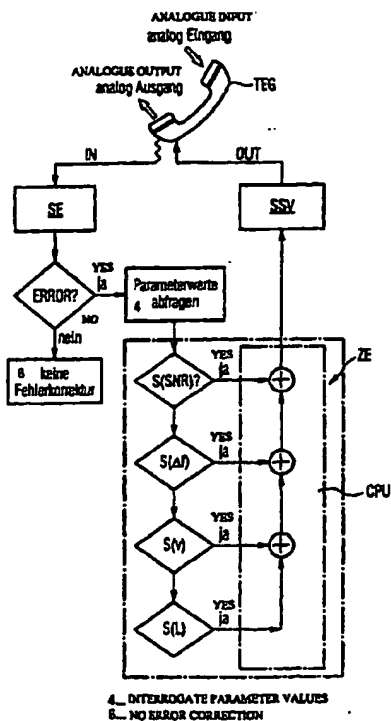
(10) Internationale Veröffentlichungsnummer
WO 01/39176 A2

- (51) Internationale Patentklassifikation⁷: G10L 15/00 (72) Erfinder; und
(21) Internationales Aktenzeichen: PCT/DE00/04121 (75) Erfinder/Anmelder (nur für US): BAUER, Josef
(22) Internationales Anmeldedatum: 22. November 2000 (22.11.2000) (74) Gemeinsamer Vertreter: SIEMENS AKTIENGE-
(25) Einreichungssprache: Deutsch (DE/DE); Winzerer Strasse 96, 80797 München (DE).
(26) Veröffentlichungssprache: Deutsch (81) Bestimmungsstaat (national): US.
(30) Angaben zur Priorität: 199 56 747.6 25. November 1999 (25.11.1999) DE (84) Bestimmungsstaaten (regional): europäisches Patent (AT,
(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,
US): SIEMENS AKTIENGESELLSCHAFT [DE/DE]; NL, PT, SE, TR).
Winzelsbacher Platz 2, 80333 München (DE).

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD AND DEVICE FOR VOICE RECOGNITION AND A TELECOMMUNICATIONS SYSTEM

(54) Bezeichnung: VERFAHREN UND VORRICHTUNG ZUR SPRACHERKENNUNG SOWIE EIN TELEKOMMUNIKATI-
ONSSYSTEM



(57) Abstract: The invention relates to a method for voice recognition. According to said method, at least one voice parameter is analysed during the voice recognition process. If said voice parameter exceeds a threshold (S), a message (OUT) is issued to the user, which is specifically designed to request observance of the value range that has been predetermined for the voice parameter. Said message prompts the user to re-input the command with a correction which has been adjusted to the voice parameter.

(57) Zusammenfassung: Beim erfindungsgemäßen Verfahren zur Spracherkennung wird während der Spracherkennung mindestens ein Sprechparameter analysiert und bei Überschreiten eines Schwellwertes (S) durch den Sprechparameter eine auf die Wiedereinhaltung des für den Sprechparameter festgelegten Wertebereich individualisierte Nachricht (OUT) an den Anwender ausgegeben, mit welcher er zur erneuten Eingabe mit an den Sprachparameter angepaßter Korrektur aufgefordert wird.

WO 01/39176 A2



Veröffentlicht:

- Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

Beschreibung

Verfahren und Vorrichtung zur Spracherkennung sowie ein Telekommunikationssystem

5

Die vorliegende Erfindung betrifft ein Verfahren zur Spracherkennung gemäß dem Oberbegriff des Patentanspruchs 1, eine Vorrichtung gemäß dem Oberbegriff des Patentanspruchs 11 sowie ein Telekommunikationssystem.

10

In dem Bereich der Mensch-Maschine-Kommunikation sind überwiegend Ansätze zur Auswertung des Informationsgehaltes der menschlichen Sprache anzutreffen, da gerade das gesprochene Wort im alltäglichen Leben der Menschen zur leichten, schnellen und sehr kompakten Übermittlung von gezielter Information einen hohen Stellenwert besitzt. Aufgrund der flächendeckenden Verfügbarkeit und des vertrauten Umgangs hat sich das Telefon als Übertragungsmedium für das gesprochene Wort im Alltagsleben durchgesetzt. Mit dem Ziel einer Erleichterung und Automatisierung einfacher Teile des Informationsaustausches zwischen Mensch und Maschine via Telefon werden beispielsweise bei der Bestellungsannahmen in Call-Centern oder im Zuge des Tele-bankings Auskunfts- und Auftragsannahmesysteme in Banken mit Verfahren und Vorrichtungen zu Spracherkennung eingesetzt.

Bisher bekannte nutzerunabhängige Verfahren und Vorrichtungen der Spracherkennung weichen oftmals von dem am Telefon vertrauten, spontanen und natürlichen Umgang der Menschen erheblich ab. Fehlfunktionen in Form von Spracherkennungsfehler sind bei bekannten Systemen häufig, da bekannte Analyseverfahren empfindlich auf Besonderheiten der jeweiligen Eingabesignale reagieren, wie z.B. die Sprechweise eines Benutzers. So steigt die Fehlerrate z.B. bei starken Hintergrundgeräuschen und auch bei sehr schnellen oder zu langsamen Sprechen bei per Telefon übertragenen Sprachsignalen stark an. Dadurch können nahezu unbrauchbare Ergebnisse geliefert werden. Zur

Behebung dieses Problems ist es bekannt, den Anwender aufzufordern nochmals deutlich zu sprechen. Es wird dann eine automatische Ansage erzeugt, die folgendermaßen lauten kann:
„Ich habe Sie nicht verstanden, bitte sprechen Sie deutlicher“.

Zur Verbesserung der Spracherkennung unter Beibehaltung eines möglichst natürlichen Sprechrhythmus in der menschlichen Sprache werden komplexe Verfahren zur gesonderten Anpassung der Maschine an jeden einzelnen Anwender vorgeschlagen, wie sie z.B. in einer Zusammenfassung in dem Buch „Anwendungsspezifische Online-Anpassung von Hidden-Markov-Modellen in automatischen Spracherkennungssystemen“ von Udo Bub, Herbert Utz Verlag, München, 1999, dargestellt sind.

Es ist die Aufgabe der vorliegenden Erfindung, ein Verfahren, eine Vorrichtung sowie ein Telekommunikationssystem zur nutzerunabhängigen Spracherkennung zu schaffen, wobei die Akzeptanz des Anwenders durch ein natürliches, dem menschlichen Gebrauch angepaßtes Verfahren bzw. durch eine Vorrichtung zur Umsetzung dieses Verfahrens bzw. ein dementsprechendes Telekommunikationssystem gesteigert wird.

Diese Aufgabe wird durch ein Verfahren mit den Merkmalen des Patentanspruchs 1, durch eine Vorrichtung gemäß Patentanspruch 11 bzw. durch ein Telekommunikationssystem gemäß Anspruch 13 gelöst.

Das Verfahren umfaßt das Analysieren eines Sprechparameters während der Spracherkennung, wobei bei Überschreitung eines Schwellwertes durch den Sprechparameter an den Anwender eine Aufforderung ausgegeben wird, die konkret auf die Wiedereinhaltung des für den Sprechparameter festgelegten Wertebereichs gerichtet ist. Während in bekannten Verfahren von dem Benutzer eine starre Anpassung an das System verlangt wird, so daß die Akzeptanz des Benutzers u.a. aufgrund eines damit verbundenen Mangels an Natürlichkeit absinkt, analysiert ein

erfindungsgemäßes Verfahren die Qualität des eingehenden Sprachsignals und fordert den Anwender mittels einer Nachricht, die anhand des Sprechparameters individualisiert ist, zu einer erneuten Spracheingabe auf. Es wird also der Anwender gezielt zu einer aktiven Anpassung seiner Sprechweise bewegt.

Definitionen von Sprechparametern der in Frage kommenden Parameter des zu verarbeitenden Sprachsignals eines Menschen bzw. Anwenders sind in den Unteransprüchen beispielhaft aufgeführt. Im Rahmen einer möglichen Umsetzung eines erfindungsgemäßen Verfahrens kann der Anwender in einer bevorzugten Ausführungsform konkret den Satz „Sprechen Sie bitte etwas leiser“ zu hören bekommen, ganz so wie bei der Unterhaltung mit einem anderen Menschen auch.

In einer Weiterbildung können für einen Sprechparameter auch mehrere Schwellwerte festgelegt werden. Bei Überschreiten der unterschiedlichen Schwellwerte wird die Aussagekraft der auszugebenden Nachricht entsprechend angepaßt. Damit ergibt sich speziell in dem vorstehend beispielhaft dargestellten Fall der Korrektur in Bezug auf die Lautstärke eine Korrekturbandbreite von „leiser“, „etwas lauter“ bis „lauter“.

Anhand des Sprechparameters kann eine Kenngröße für die Qualität des eingehenden Sprachsignals bestimmt werden, die auch als Hinweis für die Güte der Spracherkennung ausgewertet werden kann. Auch kann anhand anhaltender Schwellwertüberschreitungen ein systematischer Fehler erkannt werden. Wird z.B. auf einem Übertragungskanal eines mit einem erfindungsgemäßen Spracherkennungsvorrichtung versehenen Telekommunikationssystems derartiges festgestellt, so kann im Rahmen des beschriebenen Verfahrens eine Kanalmessung veranlaßt werden. In diesem Fall kann es erfindungsgemäß sogar vorgesehen sein, den Anwender unter Hinweise auf eine vermutete Störung zur Benutzung eines anderen Telefonendgerätes aufzufordern.

Weiter umfaßt eine erfindungsgemäße Vorrichtung zur Spracherkennung mit mindestens einer Einrichtung zum Verarbeiten digitalisierter Daten eines Sprachsignals und einer Einrichtung zur Sprachausgabe, mit

- 5 - Einrichtungen zur Analyse und Überwachung eines Sprechparameters,
- einer Einrichtung zum Feststellen einer Überschreitung des Sprechparameters und
- einer Einrichtung zur Erzeugung und Ausgabe eines in Abhän-
- 10 - gigkeit von der Überschreitung eines Schwellwert für den Sprechparameter generierten Hinweises in digitaler oder analoger Form, insbesondere einer Sprachsyntheseeinrichtung, sowie
- einer Einrichtung zur Übermittlung des Hinweises an einen
- 15 - das Sprachsignal erzeugenden Anwender.

Ein erfindungsgemäßes Telekommunikationssystem umfaßt demnach

- einer Vielzahl von Telefonendgeräten,
- Wandlereinrichtungen zur Digital/Analog- und Ana-
- 20 - log/Digital- Wandlung und Signalaufbereitung,
- einer Verbindungsleitung für jedes der Telefonendgeräte,
- Kanalbündelungs- und Kanalaufzweigungsmitteln,
- mindestens einer Vermittlungsstelle und
- eine Einrichtung zur Spracherkennung.

25

Vorteilhafte Weiterbildungen eines erfindungsgemäßen Verfahrens zur Spracherkennung bzw. einer erfindungsgemäßen Vorrichtung sind Gegenstand von Unteransprüchen.

- 30 Die vorliegende Vorrichtung zur Spracherkennung unter Anwendung eines erfindungsgemäßen Verfahrens wird im folgenden unter Bezugnahme auf die zugehörigen Zeichnungen näher erläutert. Sie zeigen:

- 35 Fig. 1 einen Prinzipschaltplan eines erfindungsgemäßen Telekommunikationssystems und

Fig. 2 ein Flußdiagramm einer Spracherkennungseinrichtung.

- In Figur 1 ist eine schematische Ansicht einer erfindungsgemäßen Vorrichtung in Form eines Telekommunikationssystems dargestellt, das nach dem erfindungsgemäßen Verfahren arbeitet. Es ist ein Telekommunikationssystem dargestellt, das eine Vielzahl von Telefonendgeräten TEG umfaßt, die Wandlereinrichtungen zur Digital/Analog- und Analog/Digital- Wandlung A/D, D/A sowie integrierter Signalaufbereitung enthalten. Die dargestellten Telefonendgeräte TEG können also digital arbeitende Geräte u.a. handelsübliche EURO-ISDN-Geräte nach dem Standard der Deutschen Telekom sein. Die Telefonendgeräte TEG sind mit einer Verbindungsleitung VL für jedes der Telefonendgeräte TEG mit Kanalbündelungs- und Kanalaufzweigungsmitteln KM, KM^{-1} verbunden, so daß die Sprachsignale S in digitalisierter Form über eine Vermittlungsstelle VS auf den Verbindungsleitungen VL zu einer Einrichtung zur Spracherkennung SEV gelangen.
- Die Einrichtung zur Spracherkennung SEV ist in eine Spracherkennungseinrichtung SE zur Verarbeitung des Eingangssignals IN und eine Sprachsyntheseeinrichtung SSV unterteilt. In der Spracherkennungseinrichtung SE (Fig. 2) wird eine Fehleruntersuchung bei der Spracherkennung durchgeführt. Bei Auftreten eines Fehlers oder bei Erreichen einer Fehlerquote wird eine Fehlermeldung FM zusammen mit bestimmten, bei der Spracherkennung ermittelten Parametern an eine Auswertelogik gesandt. Hier werden die einzelnen Parameter auf eine Überschreitung eines jeweiligen Schwellenwertes S untersucht.
- Als Parameter werden Eigenschaften eines Sprachsignals abgebildet, die einen wesentlichen Einfluß auf die Fehlerquote bei der Spracherkennung ausüben. Nachfolgend werde einige Beispiele für derartige Parameter vorgestellt:
- Ein solcher Parameter ist die Lautstärke L des Sprachsignals. Der Wert des Parameters L kann aus dem analogen wie auch dem

digitalen Sprachsignal gewonnen werden. Von der Spracherkennungseinrichtung SEV kann kein Einfluß auf die Eingangsverstärkung des Sprachsignals an einem Telefonendgerät TEG (Fig. 1), da diese Geräte mit festen Eigenschaften systembedingt unabhängig sind, und das von der nachgeschalteten Übertragungsstrecke sowie von den Anforderungen der Spracherkennungseinrichtung SEV als Empfangsgerät ausgeübt werden.

Die Ausgabe einer akustischen Nachricht an den Anwender kann durch den Einsatz gestaffelter Schwellwerte präzisiert werden. So wird bei Überschreitung der entsprechenden Schwellwerte für die Lautstärke L der Sprecher aufgefordert leiser, etwas lauter bzw. lauter zu sprechen.

Ein weiterer Parameter ist der Signal-Rausch-Abstand SNR eines Sprachsignals. Ist der Signal-Rausch-Abstand SNR in dem an der Spracherkennungseinrichtung SE anliegenden Sprachsignal IN zu gering, so kann die Sprache nicht fehlerfrei erkannt werden. Es gibt zwar einige Möglichkeiten, den Signal-Rausch-Abstand automatisch zu verbessern, wie z.B. spezielle digitale Filterverfahren, deren Filterparameterwerte dem aktuellen Fall entsprechend eingestellt werden oder auch Verfahren wie die Autokorrelation zur nachträglichen Verbesserung der Kanalübertragungseigenschaften.

Bei einer Überschreitung des Schwellwertes für den Signal-Rausch-Abstand SNR kann zunächst die Lautstärke L geprüft werden. Ist die Lautstärke gering, so wird, auch wenn der für die Lautstärke geltende Schwellwert noch nicht unterschritten ist, der Sprecher aufgefordert lauter zu sprechen. Hierdurch stellt sich ein größerer Signal-Rausch-Abstand ein. Ist der sich hierdurch ergebende Signal-Rausch-Abstand immer noch nicht ausreichend oder ist die Lautstärke L nicht gering, so liegen ungünstige Umstände vor. Z.B. kann der Sprecher in einem lauten Umfeld sprechen (bspw. Wartehallen von Bahnhöfen und Flughäfen) oder die Übertragung ist gestört. Der Sprecher

wird dann bspw. aufgefordert, von einem anderen Ort oder einem anderen Telefon zu sprechen.

Ein weiterer wichtiger Parameter ist die Sprechgeschwindigkeit v , die zu hoch oder zu niedrig liegen kann. Die Sprechgeschwindigkeit v wird beispielsweise durch eine Messung der Phoneme pro Zeit festgestellt, wobei unter dem Begriff der Phoneme die kleinste linguistische Grundeinheit einer Sprache verstanden wird, die bedeutungsunterscheidend ist. Wie ein Mensch auch kann eine Maschine zu schnell gesprochener Sprache und einer dementsprechend raschen Abfolge von Phonemen nicht mehr folgen, wodurch die Fehlerquote stark ansteigt. Insbesondere ist bekannt, daß die Erkennungsrate bei Zifferneingabe erheblich mit zunehmender Sprechgeschwindigkeit absinkt. Bei Satzerkennungsverfahren, die mehrere Wörter oder ganze Sätze auf einmal verarbeiten, macht hingegen auch eine zu geringe Sprechgeschwindigkeit Probleme, da das System nun ungewöhnlich lange auf das Eintreffen eines Sprachstücks, das es verarbeiten kann, warten muß.

Der Sprecher wird bei Überschreitung der entsprechenden Schwellwerte aufgefordert, langsamer oder schneller zu sprechen.

Als eine weitere Quelle für eine erhöhte Fehlerquote bei der Spracherkennung sind spektrale Eigenschaften des Sprachsignals anzugeben. Das durch die Übertragung per Telefon auf ein enges Frequenzband beschränkte Sprachsignal weist bei allen menschlichen Sprechern Gemeinsamkeiten auf, die bei der Spracherkennung genutzt werden können. In Abhängigkeit von den jeweils eingesetzten Aufnahmемикrophonen kann es hier schon zu Unterschieden kommen. Da in Telefonendgeräten jedoch stets qualitativ ungefähr gleichwertige Aufnahmемикrophone eingesetzt werden, ist dieser Einfluß vernachlässigbar gegenüber dem Einfluß des Winkels und des Abstandes, unter dem in das Aufnahmемикrophon gesprochen wird. Durch eine Abweichung in der Lautstärke und in den spektrale Eigenschaften des

Sprachsignals kann erkannt werden, daß ein Sprecher nicht direkt und aus kurzer Distanz in ein Aufnahmемикrophon spricht. Als Parameter wird daher eine spektrale Frequenzabweichung Δf festgelegt, dessen Wert durch eine Richtcharakteristik des Aufnahmемикrophons zusammen mit einem Einfallswinkel des Sprachsignals auf das Aufnahmемикrophon erzeugt wird.

Eine Überschreitung der Schwellwerte der spektralen Frequenzabweichung bedeutet somit, daß der Sprecher nicht das Mikrofon bzw. den Hörer eines Telefons vor seinem Mund angeordnet hat. In einem solchen Fall wird der Sprecher aufgefordert, das Mikrofon nahe an seinem Mund anzuordnen.

Die genannten Parameter stellt mithin einen Gütemaßstab für das zu erkennende Sprachsignal IN dar. In der Ausführungsform von Fig. 1 wird das in die SEV eingehende digitalisierte Sprachsignal IN nun direkt in die Spracherkennung SE eingegeben. Hier wird eine laufende Fehlerkontrolle durchgeführt. Im Fall des Auftretens von Fehlern werden die Werte der Parameter Signal-Rausch-Abstand SNR, spektrale Frequenzabweichung Δf , Sprechgeschwindigkeit v und Lautstärke L an eine Zentraleinheit ZE geliefert. Hier sind Schwellwertmeßeinrichtungen SW angeordnet, die bei Überschreitung eines jeweiligen Schwellenwertes S_i für die genannten Parameter ein eigenes Steuersignal an eine Sprachsyntheseeinrichtung SSV aussenden. Diese Schritte laufen alle in Echtzeit ab, um das Eingabeverfahren nicht zu verzögern. In der Sprachsyntheseeinrichtung SSV wird die Summe dieser Steuersignale verarbeitet, indem hieraus eine z.B. akustische Nachricht aufgebaut wird, die auch mehrteilig sein kann. Sind beispielsweise für die Auswertung der Lautstärke L mehrere Schwellwerte S_i vorgesehen, so kann in der Zentraleinheit ZE eine Quantisierung festgestellt werden. Es wird dann z.B. festgestellt, daß das Eingangssignal für eine korrekte Erkennung nur etwas zu leise ist. Die auf das dementsprechende Steuersignal ausgesandte Nachricht würde z.B. lauten: „Sprechen Sie bitte etwas lauter“.

Wie in Fig. 2 dargestellt prüft die Zentraleinheit ZE die Schwellwerte aller vorgegebenen Parameter vor jeder durch ein Fehlersignal bewirkten Sprachausgabe durch. Erst das durch
5 eine zentrale Recheneinheit CPU innerhalb der Zentraleinheit ZE zusammengestellte Summenergebnis aller Steuersignale wird an die Sprachsyntheseeinrichtung SSV und dort in eine digitale und nach der Analogwandlung hörbare Nachricht umgesetzt. Wird also von der Zentraleinheit ZE auch festgestellt, daß zu
10 schnell gesprochen wird lautet die fertige Nachricht an den Anwender: „Sprechen Sie bitte etwa lauter und langsamer“.

Im Unterschied zu Systemen nach dem Stand der Technik enthält die erfindungsgemäß ausgegebene Nachricht eine auf den kon-
15 kreten Fall bezogene, individualisierte Nachricht, die sehr zielgerichtet zu einer Verbesserung der Spracherkennung führt.

In dem vorliegenden Ausführungsbeispiel wird die Nachricht in
20 hörbarer Form an den Anwender übermittelt, nämlich als synthetisch generierter Satz (eine Mitteilung OUT). Zur Ausgabe der Mitteilung OUT wird ein digitales Signal in der Sprachsyntheseeinrichtung SSV erzeugt und über Verbindungsleitungen VL mit Kanalbündelungsmittel KM und Kanalaufzweigungsmittel
25 KM^{-1} über die Vermittlungsstelle VS zu dem entsprechenden Telefonendgerät TEG übermittelt, um nach der Analogwandlung als S Sprachsignal mit einer konkreten Nachricht den Anwender zu erreichen.

30 Die Nachricht OUT kann über die vorstehend beschriebene Vorrichtung in umgekehrter Richtung auch statt als hörbare Nachricht in sonstiger Weise verarbeitet werden. Beispielsweise kann die Nachricht an dem Telefonendgerät TEG dem Anwender angezeigt werden, z.B. an einem Bildschirmtelefon bzw. einem
35 PC mit integriertem Telefon oder einem Display.

In Abwandlung von dem in Fig.2 dargestellten Signalfluß kann die Fehleranalyse auch aus der Spracherkennung SE ausgegliedert und innerhalb der Einrichtung zur Spracherkennung SEV nebengeordnet sein. Damit kann die zur Verfügung stehende Re-

5 chenleistung der Spracherkennung SE voll auf diese eine Aufgabe verwendet werden, da die Analyse der Ergebnisse auf Fehler parallel dazu in einem kontinuierlichen Prozeß anhand der Untersuchung der vorstehend genannten Parameter auf Schwellwertüberschreitung in Schwellwertmeßeinrichtungen SW durchge-

10 führt werden kann. Bei Schwellwertüberschreitungen kann dann selektiv eine Fehleranalyse durchgeführt werden, so daß in dem beschriebenen nutzerunabhängigen System auch bei engen Schwellwerten die gesamte Belastung für ein System vertretbar bleiben. Eine weitere Möglichkeit besteht in der Vorgabe von

15 Eingangsschwellwerten, die schnell fallabhängig nachgeführt bzw. angepaßt werden. So wird insgesamt eine sichere Konvergenz auf geringe Fehlerquoten bei minimaler Anzahl von Anforderungen an den Anwender bzw. systeminternen Regelvorgängen erreicht.

Patentansprüche

1. Verfahren zur Spracherkennung,
bei dem während der Spracherkennung festgestellt wird, ob bei
5 der Spracherkennung ein Fehler auftritt, und falls ein ent-
sprechender Fehler festgestellt wird, mindestens ein Sprech-
parameter analysiert wird und
bei Über- oder Unterschreiten eines Schwellwertes (S) durch
den Sprechparameter eine auf die Wiedereinhaltung des für den
10 Sprechparameter festgelegten Wertebereichs individualisierte
Nachricht (OUT) an den Anwender ausgegeben wird, mit welcher
er zur erneuten Eingabe mit an den Sprachparameter angepaßter
Korrektur aufgefordert wird.
- 15 2. Verfahren nach Anspruch 1,
d a d u r c h g e k e n n z e i c h n e t,
daß als vorbestimmter Sprechparameter ein Signal-Rausch-
Abstand (SNR) des eingegebenen Sprachsignals (IN) überwacht
wird.
- 20 3. Verfahren nach einem der vorhergehenden Ansprüche,
d a d u r c h g e k e n n z e i c h n e t,
daß als vorbestimmter Sprechparameter die Einhaltung eines
bestimmten Bereichs des Frequenzspektrums (Δf) des eingege-
25 benen Sprachsignals (IN) überwacht wird.
4. Verfahren nach einem der vorhergehenden Ansprüche,
d a d u r c h g e k e n n z e i c h n e t,
daß als vorbestimmter Sprechparameter die Sprechgeschwindig-
30 keit (v) innerhalb des eingegebenen Sprachsignals (IN) über-
wacht wird.
5. Verfahren nach einem der vorhergehenden Ansprüche,
d a d u r c h g e k e n n z e i c h n e t,
35 daß als vorbestimmter Sprechparameter die Lautstärke (L) des
eingegebenen Sprachsignals überwacht wird.

6. Verfahren nach einem der vorhergehenden Ansprüche 3-5, dadurch gekennzeichnet, daß mehrere Sprechparameter (SNR, Δf , v. L) gleichzeitig überwacht werden.
- 5
7. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Überwachung jedes der Sprechparameter (SNR, Δf , v, L) in dem digitalisierten Sprachsignal (IN) vorgenommen wird.
- 10
8. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Aufforderung (OUT) an den Anwender akustisch in Form eines gesprochenen kurzen Satzes über eine Sprachsyntheseeinrichtung (SSV) ausgegeben wird.
- 15
9. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß für den Sprechparameter (SNR, Δf , v, L) zur weiteren
- 20
- Präzisierung der an den Anwender ergehenden Nachricht (OUT) mehr als ein Schwellwert (S) vorgesehen ist.
10. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet,
- 25
- daß der Sprechparameter in Echtzeit aus dem eingegebenen und verarbeitenden Sprachsignal (IN) bestimmt wird.
11. Vorrichtung zur Spracherkennung mit mindestens einer Einrichtung zum Verarbeiten digitalisierter Daten eines Sprachsignals insbesondere nach einen der Ansprüche 1-10 und einer Sprachausgabe, mit
- 30
- einer Einrichtung zur Spracherkennung mit Analyse und Überwachung eines Sprechparameters,
 - einer Einrichtung zum Feststellen einer Überschreitung des
 - 35 Sprechparameters und
 - einer Einrichtung zur Erzeugung und Ausgabe eines in Abhängigkeit von der Überschreitung eines Schwellwert für den

13

Sprechparameter (SNR, Δf , v , L) generierten Hinweises in digitaler oder analoger Form, insbesondere einer Sprachsyntheseeinrichtung (SSV) sowie

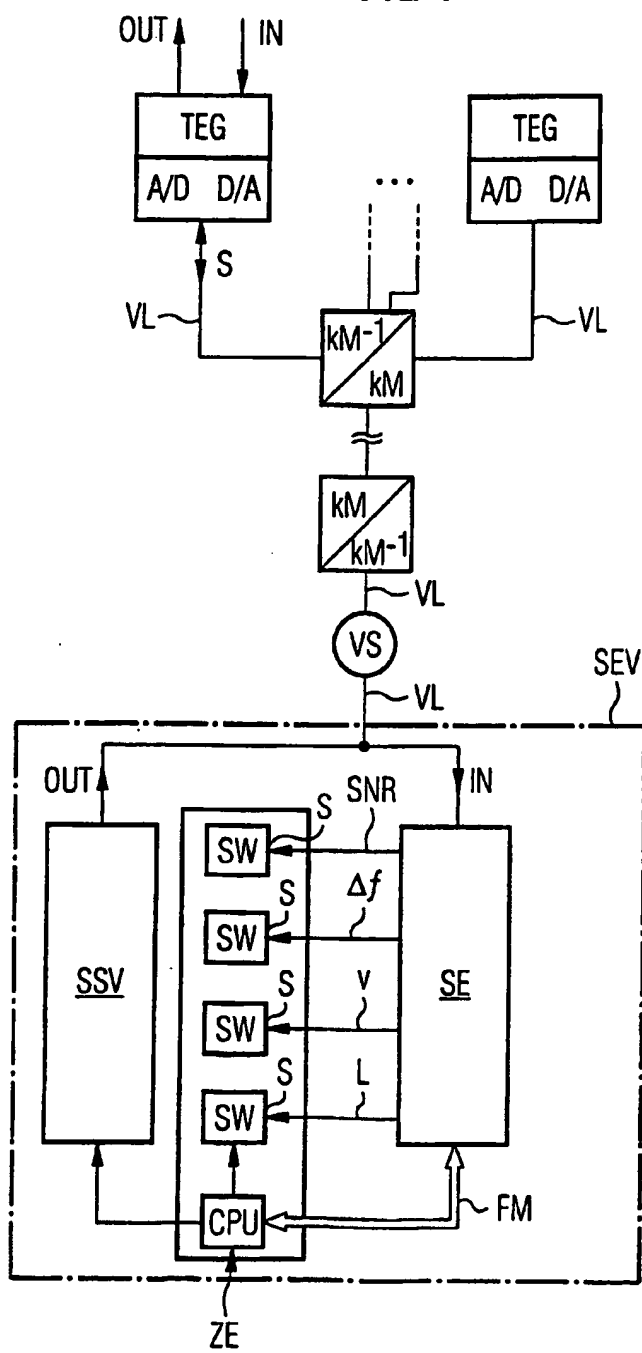
- 5 - einer Einrichtung zur Übermittlung des Hinweises an einen das Sprachsignal (IN) erzeugenden Anwender.

12. Vorrichtung nach Anspruch 11,
d a d u r c h g e k e n n z e i c h n e t,
daß als Einrichtung zur Übermittlung des Hinweises an einen
10 das Sprachsignal erzeugenden Anwender ein digitales oder analoges Telefonsystem vorgesehen ist.

13. Telekommunikationssystem mit Spracherkennung mit
- 15 - einer Vielzahl von Telefonendgeräten (TEG),
- 15 - Wandlereinrichtungen zur Digital/Analog- und Analog/Digital- Wandlung (D/A, A/D) und Signalaufbereitung,
- 15 - einer Verbindungsleitung (VL) für jedes der Telefonendgeräte (TEG),
- 15 - Kanalbündelungs- und Kanalaufzweigungsmitteln (KM , KM^{-1}) ,
- 20 - mindestens einer Vermittlungsstelle (VS) und
- 20 - mindestens einer Vorrichtung nach Anspruch 11 oder 12.

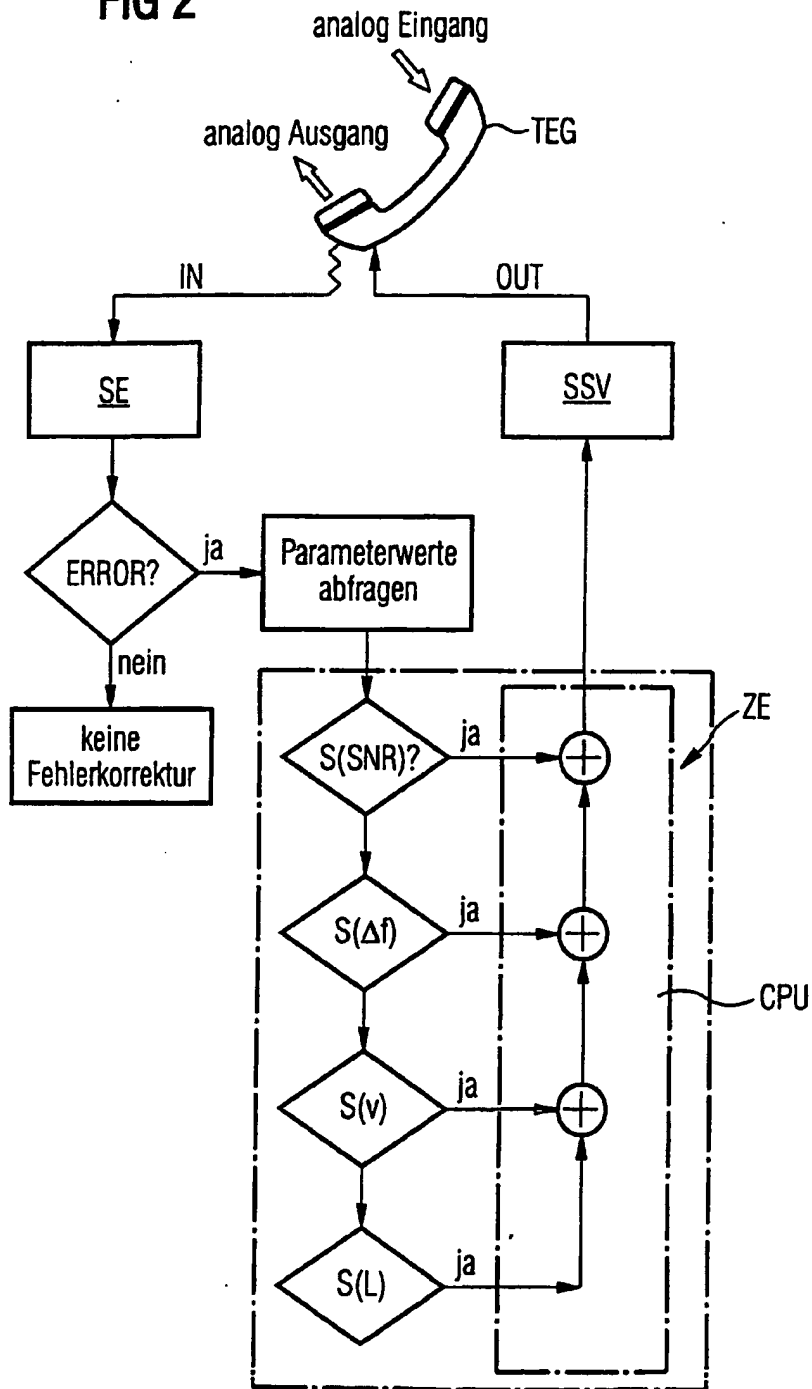
1/2

FIG 1



2/2

FIG 2



(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
31. Mai 2001 (31.05.2001)

PCT

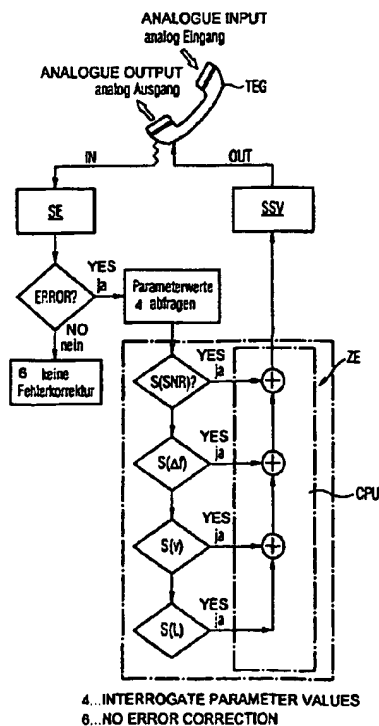
(10) Internationale Veröffentlichungsnummer
WO 01/039176 A3

- (51) Internationale Patentklassifikation⁷: **G10L 15/22, 15/20**
- (21) Internationales Aktenzeichen: **PCT/DE00/04121**
- (22) Internationales Anmeldedatum:
22. November 2000 (22.11.2000)
- (25) Einreichungssprache: **Deutsch**
- (26) Veröffentlichungssprache: **Deutsch**
- (30) Angaben zur Priorität:
199 56 747.6 25. November 1999 (25.11.1999) **DE**
- (71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): **SIEMENS AKTIENGESELLSCHAFT [DE/DE]**, Wiltshbacher Platz 2, 80333 München (DE).
- (72) Erfinder; und
(75) Erfinder/Anmelder (nur für US): **BAUER, Josef [DE/DE]**; Winzerer Strasse 96, 80797 München (DE).
- (74) Gemeinsamer Vertreter: **SIEMENS AKTIENGESELLSCHAFT**; Postfach 22 16 34, 80506 München (DE).
- (81) Bestimmungsstaat (national): **US.**
- (84) Bestimmungsstaaten (regional): europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

[Fortsetzung auf der nächsten Seite]

(54) Title: **METHOD AND DEVICE FOR VOICE RECOGNITION AND A TELECOMMUNICATIONS SYSTEM**

(54) Bezeichnung: **VERFAHREN UND VORRICHTUNG ZUR SPRACHERKENNUNG SOWIE EIN TELEKOMMUNIKATIONSSYSTEM**



(57) Abstract: The invention relates to a method for voice recognition. According to said method, at least one voice parameter is analysed during the voice recognition process. If said voice parameter exceeds a threshold (S), a message (OUT) is issued to the user, which is specifically designed to request observance of the value range that has been predetermined for the voice parameter. Said message prompts the user to re-input the command with a correction which has been adjusted to the voice parameter.

(57) Zusammenfassung: Beim erfindungsgemäßen Verfahren zur Spracherkennung wird während der Spracherkennung mindestens ein Sprechparameter analysiert und bei Überschreiten eines Schwellenwertes (S) durch den Sprechparameter eine auf die Wiedereinhaltung des für den Sprechparameter festgelegten Wertebereichs individualisierte Nachricht (OUT) an den Anwender ausgegeben, mit welcher er zur erneuten Eingabe mit an den Sprachparameter angepaßter Korrektur aufgefordert wird.

WO 01/039176 A3

Veröffentlicht:

— mit internationalem Recherchenbericht

(88) Veröffentlichungsdatum des internationalen

Recherchenberichts:

26. September 2002

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/DE 00/04121

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L15/22 G10L15/20

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G10L H04M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	EP 1 081 683 A (PHILIPS CORP INTELLECTUAL PTY ;KONINKL PHILIPS ELECTRONICS NV (NL)) 7 March 2001 (2001-03-07) column 1, line 1 - line 45 column 2, line 30 - line 34	1,4,5,8, 10-13
X	WO 87 07460 A (INNOVATIVE DEVICES) 3 December 1987 (1987-12-03) page 4 page 16 -/-	1,2,5,8, 10-13

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

A document member of the same patent family

Date of the actual completion of the international search

11 July 2001

Date of mailing of the international search report

26/07/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Ramos Sánchez, U

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/DE 00/04121

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NAIK D: "USING TALKER LOCATION TO DETECT SPURIOUS UTTERANCES IN DESKTOP COMMAND AND CONTROL" IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP),US,LOS ALAMITOS,CA: IEEE COMP. SOC. PRESS, 21 April 1997 (1997-04-21), pages 3265-3268, XP000788057 ISBN: 0-8186-7920-4 the whole document	1,2,11, 13
A	DE 195 15 123 C (ORGA KARTENSYSTEME GMBH) 22 August 1996 (1996-08-22) column 1, line 1 - line 18 column 5, line 40 - line 47 claims 1,9,10	1,3,11, 13
A	GB 2 102 171 A (PARKHOUSE JOHN GRAHAM;TURRELL RONALD PETER) 26 January 1983 (1983-01-26) column 2, line 84 - line 98	1,4,11, 13

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/DE 00/04121

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 1081683	A	07-03-2001	DE 19941227 A	08-03-2001
WO 8707460	A	03-12-1987	AU 7488287 A	22-12-1987
DE 19515123	C	22-08-1996	NONE	
GB 2102171	A	26-01-1983	US 4472833 A	18-09-1984

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/DE 00/04121

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES IPK 7 G10L15/22 G10L15/20		
Nach der internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK		
B. RECHERCHIERTE GEBIETE Recherchierte Mindestprüfstoff (Klassifikationssystem und Klassifikationsymbole) IPK 7 G10L H04M		
Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen		
Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe) EPO-Internal		
C. ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
E	EP 1 081 683 A (PHILIPS CORP INTELLECTUAL PTY ;KONINKL PHILIPS ELECTRONICS NV (NL)) 7. März 2001 (2001-03-07) Spalte 1, Zeile 1 - Zeile 45 Spalte 2, Zeile 30 - Zeile 34	1,4,5,8, 10-13
X	WO 87 07460 A (INNOVATIVE DEVICES) 3. Dezember 1987 (1987-12-03) Seite 4 Seite 16	1,2,5,8, 10-13
--- -/--		
<input checked="" type="checkbox"/> Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen		
<input checked="" type="checkbox"/> Siehe Anhang Patentfamilie		
* Besondere Kategorien von angegebenen Veröffentlichungen : *A* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist *E* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist *L* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt) *O* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht *P* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist *T* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist *X* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderscher Tätigkeit beruhend betrachtet werden *Y* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderscher Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist *Z* Veröffentlichung, die Mitglied derselben Patentfamilie ist		
Datum des Abschlusses der internationalen Recherche 11. Juli 2001		Absendedatum des internationalen Recherchenberichts 26/07/2001
Name und Postanschrift der internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016		Bevollmächtigter Bediensteter Ramos Sánchez, U

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/DE 00/04121

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	<p>NAIK D: "USING TALKER LOCATION TO DETECT SPURIOUS UTTERANCES IN DESKTOP COMMAND AND CONTROL"</p> <p>IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP), US, LOS ALAMITOS, CA: IEEE COMP. SOC. PRESS,</p> <p>21. April 1997 (1997-04-21), Seiten 3265-3268, XP000788057</p> <p>ISBN: 0-8186-7920-4</p> <p>das ganze Dokument</p>	1,2,11,13
A	<p>DE 195 15 123 C (ORGA KARTENSYSTEME GMBH)</p> <p>22. August 1996 (1996-08-22)</p> <p>Spalte 1, Zeile 1 - Zeile 18</p> <p>Spalte 5, Zeile 40 - Zeile 47</p> <p>Ansprüche 1,9,10</p>	1,3,11,13
A	<p>GB 2 102 171 A (PARKHOUSE JOHN GRAHAM;TURRELL RONALD PETER)</p> <p>26. Januar 1983 (1983-01-26)</p> <p>Spalte 2, Zeile 84 - Zeile 98</p>	1,4,11,13

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/DE 00/04121

Im Recherchenbericht angeführtes Patentdokument		Datum der Veröffentlichung	Mitglied(er) der Patentfamilie		Datum der Veröffentlichung
EP 1081683	A	07-03-2001	DE	19941227 A	08-03-2001
WO 8707460	A	03-12-1987	AU	7488287 A	22-12-1987
DE 19515123	C	22-08-1996	KEINE		
GB 2102171	A	26-01-1983	US	4472833 A	18-09-1984